

Econometrics



David Preinerstorfer

Institute for Statistics and Mathematics
Dept. of Finance, Accounting and Statistics

Spring 2026



- Organization
- Introduction: Basic Concepts of Econometric Modeling

Title:	Econometrics (4355)
Instructor:	Prof. David Preinerstorfer, PhD
Contact details:	david.preinerstorfer@wu.ac.at

- Lectures: Please check syllabus for the exact times and locations.
- Material: Lecture slides from Zehra Eksi-Altay (based on older lecture notes of Alois Geyer and slides of Sylvia Frühwirth-Schnatter) are available on Canvas.
- Tutorials: by Filip Kindermann.
- Attendance: mandatory (attend at least 80% of all lectures)

Assessment and Date of Exams

- Active participation in class (5%) in class (nametags).
- Homework assignments (30%)
 - Submission as a group of 3 (please get assigned to a group in Canvas asap)
- Two written exams (65%):
 - Midterm (30%): 19.05.2026, 14:00–16:00 (60 mins exam). 10 short questions (true/false + brief but sufficient justification, definitions, properties, derivations, interpretation of R output), closed book.¹
 - Final Exam (35%): 24.06.2026, 13:30–15:30 (90 mins exam). 10 short questions (style of midterm) + 2 long questions, closed book.²

Content of the exams: everything covered until exam date (i.e., endterm test covers all material); bring a non-programmable calculator, study/participate continuously!

¹Example for a short question: Is it true that the OLS residuals sum up to 0?

²Example for a long question: State the assumptions we were working with in the multiple regression model, explain which minimization problem is solved by the OLS estimator, and formally derive this estimator as the solution to that optimization problem; show formally that the estimator is unbiased, and derive its conditional covariance matrix.

- Part I: Regression Analysis
- Part II: Time Series Analysis

- Organization
- Introduction: Basic Concepts of Econometric Modeling

Question: What is Econometrics? What is Applied Econometrics about?

Question: What is Econometrics? What is Applied Econometrics about?

Question: What do you know about deterministic vs. stochastic models?

Question: What is Econometrics? What is Applied Econometrics about?

Question: What do you know about deterministic vs. stochastic models?

Question: What different data types do you know? How does the data type affect the method choice?

Question: What is Econometrics? What is Applied Econometrics about?

Question: What do you know about deterministic vs. stochastic models?

Question: What different data types do you know? How does the data type affect the method choice?

Question: What's the benefit of experimental data? Is this data type frequently available in finance/economics?

Question: What is Econometrics? What is Applied Econometrics about?

Question: What do you know about deterministic vs. stochastic models?

Question: What different data types do you know? How does the data type affect the method choice?

Question: What's the benefit of experimental data? Is this data type frequently available in finance/economics?

Question: What is the simple linear regression model? How do we estimate it? What are that model's shortcomings?

Econometrics deals with *learning* about an economic phenomenon *from data* (e.g., status of the economy, influence of product attributes, volatility on financial markets)

- **Econometric model:** description of the phenomenon involving quantities that are observable
- **Data** are collected for the observable variables
- **Econometric inference:** draw conclusions from the data about the phenomenon of interest

Example: relationship between price and demand

- Description of the phenomenon involving quantities that are observable
- Economic model: simplified description of the process behind the data based on a deterministic, mathematical model;
- Sometimes the deterministic model is based on an economic theories
- Stochastic model rather than a deterministic model because of simplification.

Deterministic Economic Model

Exact quantitative relationship between the variables of interest is assumed to be known

Deterministic Relationship between Demand and Price

$$D = f(p),$$

where D is the demand and p is the price.

Linear model:

$$D = \beta_0 + \beta_1 p$$

Non-linear model:

$$D = \beta_0 p^{\beta_1}$$

Exact quantitative relationship between the variables of interest is NOT known, but disturbed by a (stochastic) error term

Stochastic Relationship between Demand and Price

$$D = f(p, \varepsilon)$$

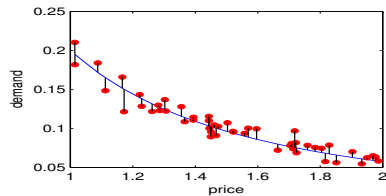
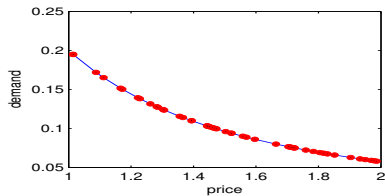
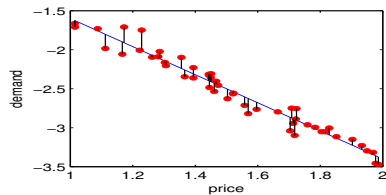
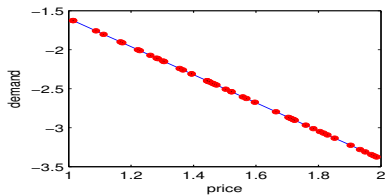
where D is the demand, p is the price, and ε is an unobservable error.

Linear model:

$$D = \beta_0 + \beta_1 p + \varepsilon$$

Non-linear model:

$$D = \beta_0 p^{\beta_1} \varepsilon$$



Where does the error come from?

- ε aggregates variables, that are not included into the model because
 - their influence is not known a priori
 - these variables are unobservable or difficult to quantify
- ε aggregates measurement errors which are caused by quantifying economic variables
- ε captures the unpredictable randomness in the left hand side variable of the model

Relationship between Demand and Price

Estimate β_1 and β_2 from the linear model:

$$D = \beta_1 + \beta_2 p + \varepsilon$$

or from the non-linear model:

$$D = \beta_1 p^{\beta_2} \varepsilon$$

from data. For the second model β_2 is the price elasticity.

Econometric inference is, in general, concerned with drawing conclusions from observed data about quantities that are not directly observable, e.g.:

- model parameters of economic interest, e.g., price elasticities
- hypothesis about these parameters
- predict the expected demand for a particular price

Need to deal with uncertainty in an appropriate way.

- Model formulation
- Model estimation
- Econometric inference: parameter estimation, hypothesis testing, forecasting/prediction
- Model specification and model selection
- Model diagnostics

- A software package is needed for practical econometric inference
- Implementations in R

Experimental data: data obtained through a designed experiment (medicine, life sciences, ...) - rare situation in economics/finance.

In economics/finance we mostly deal with **observational (non-experimental) data**

Typical data structures:

- Cross-sectional data
- Time series data

Cross-sectional data

- we are interested in variables (Y, X) (e.g. relationship between demand D and price P) or a set of variables $(Y, X_1, \dots, X_K)$
- we are observing these variables simultaneously for N subjects drawn randomly from a population (e.g., for various individual, firm, supermarkets, countries)

Typically, cross sectional data are indexed as follows:

$$(y_i, x_i), \quad (y_i, x_{1i}, \dots, x_{Ki}), \quad i = 1, \dots, N$$

If the data set is not a random sample, there is a sample-selection problem.

- we are interested in a single variable Y (e.g., the return of a financial asset);
- we are observing this variable over time (e.g., every day, every month)
- data cannot be regarded as random sample; it is important to account for trends and seasonality

Typically, time series data are indexed as follows:

$$y_t, \quad t = 1, \dots, T$$

- Pooled cross-section: Random cross sections can be pooled and treated similar to normal cross section, accounting for differences over time.
- Panel data or longitudinal data: The same (random) individual observations Y_i is followed over time, i.e., we have a time series for each cross-section unit.

Typically, panel data are indexed as follows:

$$y_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T$$

The Simple Regression Model

Bivariate cross-sectional data (i.e., $K = 1$):

- We are interested in a dependent (left-hand side, explained, response) variable Y , which is supposed to depend on an explanatory (right-hand sided, independent, control, predictor) variable X
- Data: we are observing the pair of variables (Y, X) for N subjects drawn randomly from a population (e.g., for various supermarkets, for various individuals):
 $(y_i, x_i), i = 1, \dots, N$

Model formulation

The simple linear regression model describes the dependence between the variables X and Y as:

$$Y = \beta_0 + \beta_1 X + \varepsilon. \quad (1)$$

The parameters β_0 and β_1 need to be estimated:

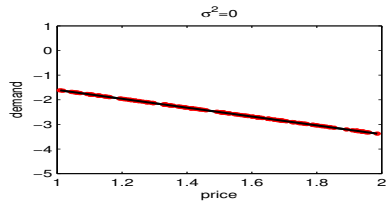
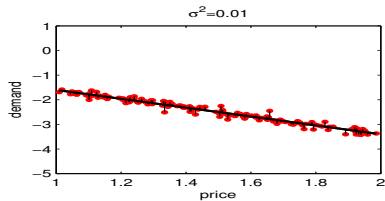
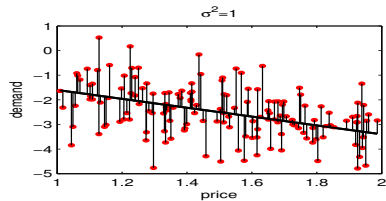
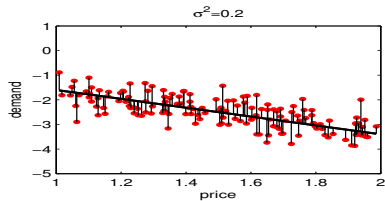
- β_0 is referred to as the constant or intercept
- β_1 is referred to as slope parameter.

- Simulate data from a simple regression model with $\beta_0 = 0.2$ and $\beta_1 = -1.8$:

$$Y = 0.2 - 1.8X + \varepsilon,$$

- Specification of the error term:

$$\varepsilon \sim \text{Normal}(0, \sigma^2)$$



Understanding the parameters

Expected value of Y , given $X = x$:

$$E(Y \mid X = x) = \beta_0 + \beta_1 x$$

Expected value of Y , if the predictor X is changed by 1:

$$E(Y \mid X = x + 1) = \beta_0 + \beta_1(x + 1).$$

Thus β_1 is the expected absolute change ΔY of the response variable Y , if the predictor X is increased by 1:

$$E(\Delta Y \mid \Delta X = 1) = E(Y \mid X = x + 1) - E(Y \mid X = x) = \beta_1.$$

The population parameters β_0 and β_1 are estimated from a sample. The parameters estimates (coefficients) are denoted by $\hat{\beta}_0$ and $\hat{\beta}_1$. Let $(y_i, x_i), i = 1, \dots, N$, denote a random sample of size N from the population. Hence, for each i :

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i. \quad (2)$$

- *Estimation problem*: how to choose the unknown parameters β_0 and β_1 ?
- The commonly used method to estimate the parameters in a simple regression model is *ordinary least square (OLS) estimation*.

- For each observation x_i , the prediction $\hat{y}_i$ of y_i depends on (β_0, β_1) :

$$\hat{y}_i(\beta_0, \beta_1) = \beta_0 + \beta_1 x_i. \quad (3)$$

- For each observation x_i define the regression residuals (prediction error) $\varepsilon_i(\beta_0, \beta_1)$ as:

$$\varepsilon_i(\beta_0, \beta_1) = y_i - \hat{y}_i(\beta_0, \beta_1) = y_i - (\beta_0 + \beta_1 x_i). \quad (4)$$

- For each parameter value (β_0, β_1) , an overall measure of fit is obtained by aggregating these prediction errors.

- The sum of squared residuals (SSR):

$$\text{SSR} = \sum_{i=1}^N \varepsilon_i(\beta_0, \beta_1)^2 = \sum_{i=1}^N (y_i - \hat{y}_i(\beta_0, \beta_1))^2 = \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i)^2. \quad (5)$$

- The OLS-estimator $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ is the parameter that minimizes the sum of squared residuals.

How to compute the OLS Estimator?

Simple regression model:

$$\hat{\beta}_1 = \frac{s_y}{s_x} r_{xy}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad (6)$$

$\bar{x}$ mean of $x_1, \dots, x_N$, $\bar{y}$ mean of $y_1, \dots, y_N$

s_x standard deviation of $x_1, \dots, x_N$, s_y standard deviation of $y_1, \dots, y_N$

r_{xy} correlation coefficient

The only requirement is that we have sample variation in X ($s_x^2 > 0$).

The OLS estimator is obtained as solution to the following minimization problem:

$$\min_{\beta_0, \beta_1} \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i)^2$$

The first-order conditions are:

$$-2 \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i) = 0 \Leftrightarrow \sum_{i=1}^N \varepsilon_i(\beta_0, \beta_1) = 0, \quad (7)$$

$$-2 \sum_{i=1}^N x_i (y_i - \beta_0 - \beta_1 x_i) = 0 \Leftrightarrow \sum_{i=1}^N x_i \varepsilon_i(\beta_0, \beta_1) = 0. \quad (8)$$

The first-order conditions are:

$$\sum_{i=1}^N \varepsilon_i(\beta_0, \beta_1) = 0, \quad (9)$$

$$\sum_{i=1}^N x_i \varepsilon_i(\beta_0, \beta_1) = 0. \quad (10)$$

where $\varepsilon_i(\beta_0, \beta_1) = (y_i - \beta_0 - \beta_1 x_i)$ are the residuals.

From (9) we have:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}. \quad (11)$$

Substituting $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ into (10) and solving for $\hat{\beta}_1$ we obtain, provided that $\sum_{i=1}^N (x_i - \bar{x})^2 > 0$ (or $s_x^2 > 0$):

$$\hat{\beta}_1 = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2} = \frac{s_y}{s_x} r_{xy}.$$

Algebraic properties of the OLS estimator

- The regression line passes through the sample midpoint.
- The sum (average) of the OLS residuals $\hat{\varepsilon}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$ is equal to zero. Follows from (9):

$$\frac{1}{N} \sum_{i=1}^N \hat{\varepsilon}_i = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0.$$

- The sample covariance between the regressor and the OLS residuals is zero. Follows from (10):

$$\frac{1}{N} \sum_{i=1}^N x_i \hat{\varepsilon}_i = \frac{1}{N} \sum_{i=1}^N x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0.$$

Question: What properties does the OLS estimator have in the simple regression model?

Question: What properties does the OLS estimator have in the simple regression model?

Question: Which assumptions are required for those properties to hold?

Question: What properties does the OLS estimator have in the simple regression model?

Question: Which assumptions are required for those properties to hold?

Question: What mathematical tools do we need to establish those properties?

Part I

Regression Analysis

- 1.1 The Multiple Regression Model
- 1.2 Finite sample properties of OLS
- 1.6 Specifications
- 1.7 Regression diagnostics
- 1.8 Generalized least squares

1.1.The Multiple Regression Model

Cross-sectional data:

- We are interested in a dependent (left-hand side, explained, response) variable Y , which is supposed to depend on K explanatory (right-hand sided, independent, control, predictor) variables $X_1, \dots, X_K$
- Examples: wage is a response and education, gender, and experience are predictor variables
- We are observing these variables for N subjects drawn randomly from a population (e.g., for various supermarkets, for various individuals): $(y_i, x_{1,i}, \dots, x_{K,i}), i = 1, \dots, N$

The multiple regression model describes the relation between the response variable Y and the predictor variables $X_1, \dots, X_K$ through a linear assumption:

Assumption AL:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_K X_K + \varepsilon, \quad (12)$$

$\beta_0, \beta_1, \dots, \beta_K$ are unknown parameters.

(12) is a linear function:

- in the parameters $\beta_0, \beta_1, \dots, \beta_K \Rightarrow$ important for „easy” OLS estimation
- in the predictor variables $X_1, \dots, X_K \Rightarrow$ important for the correct interpretation of the parameters

1.1.1 OLS-Estimation

Let $(y_i, x_{1,i}, \dots, x_{K,i}), i = 1, \dots, N$ denote a random sample of size N from the population. Hence, for each i :

$$y_i = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_K x_{K,i} + \varepsilon_i. \quad (13)$$

The population parameters $\beta_0, \beta_1, \dots, \beta_K$ are estimated from this sample. The parameters estimates (coefficients) are denoted by $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_K$. We will use the following vector notation:

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_K)', \quad \hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_K)'.$$

The commonly used method to estimate the parameters in a multiple regression model is, again, *OLS estimation*:

- For each observation y_i , the prediction $\hat{y}_i(\boldsymbol{\beta})$ of y_i depends on $\boldsymbol{\beta} = (\beta_0, \dots, \beta_K)$:

$$\hat{y}_i(\boldsymbol{\beta}) = \beta_0 + \beta_1 x_i. \quad (14)$$

$$\hat{y}_i = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_K x_{K,i}.$$

- For each y_i , define the regression residuals (prediction error) $\varepsilon_i(\boldsymbol{\beta})$ as:

$$\varepsilon_i(\boldsymbol{\beta}) = y_i - \hat{y}_i(\boldsymbol{\beta}) = y_i - (\beta_0 + \beta_1 x_{1,i} + \dots + \beta_K x_{K,i}).$$

OLS-Estimation for the Multiple Regression Model

- For each parameter value β , an overall measure of fit is obtained by aggregating these prediction errors.
- The sum of squared residuals (SSR):

$$\text{SSR} = \sum_{i=1}^N \varepsilon_i(\beta)^2 = \sum_{i=1}^N (y_i - \hat{y}_i(\beta_0, \beta_1))^2 = \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_{1,i} - \dots - \beta_K x_{K,i})^2. \quad (15)$$

- The OLS-estimator $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_K)$ is the parameter that minimizes the sum of squared residuals.

Matrix notation of the multiple regression model

Matrix notation for the observed data:

$$\mathbf{X} = \begin{pmatrix} 1 & x_{1,1} & \vdots & x_{K,1} \\ 1 & x_{1,2} & \vdots & x_{K,2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1,N-1} & \vdots & x_{K,N-1} \\ 1 & x_{1,N} & \vdots & x_{K,N} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N-1} \\ y_N \end{pmatrix}.$$

$\mathbf{X}$ is a $N \times (K + 1)$ -matrix, having N rows and $K + 1$ columns, containing in each row i the predictor values $(1 \ x_{1,i} \ \dots \ x_{K,i})$. $\mathbf{y}$ is $N \times 1$ -vector.

In matrix notation, the N equations given in (13) for $i = 1, \dots, N$, may be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{N-1} \\ \varepsilon_N \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_K \end{pmatrix}.$$

A basic assumption

The matrix $\mathbf{X}'\mathbf{X}$ is a quadratic matrix with $(K + 1)$ rows and columns.

Assumption AR (rank assumption):

We assume that $\mathbf{X}$ has full rank equal to $(K + 1)$ (i.e., the columns of $\mathbf{X}$ are linearly independent).

If $\mathbf{X}$ has full rank, then $\mathbf{X}'\mathbf{X}$ is positive definite and the inverse matrix $(\mathbf{X}'\mathbf{X})^{-1}$ of $\mathbf{X}'\mathbf{X}$ exists.

OLS Estimator

Under assumption AR, the OLS estimator $\hat{\beta}$ has an explicit form, depending on $\mathbf{X}$ and the vector $\mathbf{y}$:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (16)$$

Solving the optimization problem:

- Take the first partial derivative of (15) with respect to each parameter β_k , $k = 0, \dots, K$.
- This yields a system of $K + 1$ linear equations in $\beta_0, \dots, \beta_K$, which has a unique solution under Assumption AR.

Proof.

Take first derivatives of $SSR(\beta)$:

$$SSR(\beta) = \sum_{i=1}^N \varepsilon_i(\beta)^2 = \varepsilon' \varepsilon = (\mathbf{y} - \mathbf{X}\beta)' (\mathbf{y} - \mathbf{X}\beta) = \mathbf{y}' \mathbf{y} - 2\beta' \mathbf{X}' \mathbf{y} + \beta' \mathbf{X}' \mathbf{X} \beta$$
$$\frac{\partial SSR(\beta)}{\partial \beta} = -2\mathbf{X}' \mathbf{y} + 2\mathbf{X}' \mathbf{X} \beta = 0. \quad (17)$$

(16) results, if Assumption AR holds. The solution is a minimum since

$$\frac{\partial^2 SSR(\beta)}{\partial \beta^2} = 2\mathbf{X}' \mathbf{X}$$

is positive definite by assumption AR.

Necessary conditions for $\mathbf{X}'\mathbf{X}$ being invertible:

- We have to observe sample variation for each predictor X_k ; i.e., the sample variances of $x_{k,1}, \dots, x_{k,N}$ is positive for all $k = 1, \dots, K$.
- Furthermore, no exact linear relation between any predictors X_k and X_l should be present; i.e., the empirical correlation coefficient of all pairwise data sets $(x_{k,i}, x_{l,i})$, $i = 1, \dots, N$ is different from 1 and -1.

Perfect Multicollinearity

A sufficient assumptions about the predictors $X_1, \dots, X_K$ in a multiple regression model is the following:

- The predictors $X_1, \dots, X_K$ are not linearly dependent, i.e., no predictor X_j may be expressed as a linear function of the remaining predictors $X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_K$.

If this assumption is violated, then the OLS estimator does not exist, as the matrix $\mathbf{X}'\mathbf{X}$ is not invertible. There are infinitely many parameters values β having the same minimal sum of squared residuals, defined in (15). The parameters in the regression model are not identified.

1.1.2 Implications

Consider the estimated regression model under OLS estimation:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \dots + \hat{\beta}_K x_{K,i} + \hat{\varepsilon}_i = \hat{y}_i + \hat{\varepsilon}_i,$$

where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \dots + \hat{\beta}_K x_{K,i}$ is called the fitted value. $\hat{\varepsilon}_i$ is called the OLS residual. OLS residuals are useful:

- to estimate the variance σ^2 of the error term;
- to quantify the quality of the fitted regression model;
- for residual diagnostics

OLS residuals as proxies for the error

Compare the underlying regression model

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_K X_K + \varepsilon, \quad (18)$$

with the estimated model for $i = 1, \dots, N$:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \dots + \hat{\beta}_K x_{K,i} + \hat{\varepsilon}_i.$$

- The OLS residuals $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_N$ may be considered as a “sample” of the unobservable error ε .
- Use the OLS residuals $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_N$ to estimate the unobservable error $\sigma^2 = V(\varepsilon)$.

By condition (17), the OLS estimates satisfy the normal equation:

$$\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{X}'\hat{\boldsymbol{\varepsilon}} = 0.$$

Hence, the OLS residuals $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_N$ obey $K + 1$ linear equations and have the following algebraic properties:

Algebraic properties of the OLS estimator

- Since the first column of $\mathbf{X}$ is a column of ones, OLS estimation has the following implications: the sum (average) of the OLS residuals $\hat{\varepsilon}_i$ is equal to zero:

$$\frac{1}{N} \sum_{i=1}^N \hat{\varepsilon}_i = 0. \quad (19)$$

- The sample covariance between $x_{k,i}$ and $\hat{\varepsilon}_i$ is zero:

$$\frac{1}{N} \sum_{i=1}^N x_{k,i} \hat{\varepsilon}_i = 0, \quad \forall k = 1, \dots, K. \quad (20)$$

- The fitted values and the residuals are orthogonal:

$$\hat{\mathbf{y}}' \hat{\varepsilon} = \hat{\boldsymbol{\beta}}' \mathbf{X}' \hat{\varepsilon} = 0.$$

- The mean of the fitted values is equal to the sample mean $\bar{y}$:

$$\frac{1}{N} \sum_{i=1}^N \hat{y}_i = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{\varepsilon}_i) = \frac{1}{N} \sum_{i=1}^N y_i - \frac{1}{N} \sum_{i=1}^N \hat{\varepsilon}_i = \bar{y}.$$

- The fitted value is equal to the sample mean $\bar{y}$, if the regression equation is evaluated for the means of $\mathbf{X}$.

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N x_{k,i} \hat{\beta} = \bar{y}.$$

The OLS-estimator $\hat{\beta}$ is the parameter vector that minimizes the sum of squared residuals, defined in (15). Evidently, the minimal sum of squared residuals is equal to the sum of squares residuals, which could also be used to evaluate the model fit. How well does the multiple regression model (18) explain the variation in Y ? Compare it with the following simple model *without any predictors*:

$$Y = \beta_0 + \tilde{\varepsilon}. \quad (21)$$

The OLS estimator of β_0 minimizes the following sum of squared residuals:

$$\sum_{i=1}^N (y_i - \beta_0)^2$$

and is given by $\hat{\beta}_0 = \bar{y}$.

Coefficient of Determination

The minimal sum is equal to the total variation $SST = N s_y^2$:

$$SST = \sum_{i=1}^N (y_i - \bar{y})^2.$$

Is it possible to reduce the sum of squared residuals SST of the simple model (21) by including the predictor variables $X_1, \dots, X_K$ as in (18)? The minimal sum of squared residuals SSR of the multiple regression model (18) is always smaller than the minimal sum of squared residuals SST of the simple model (21):

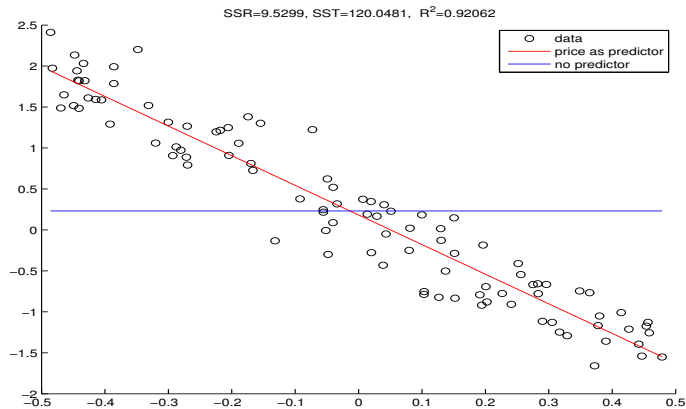
$$SSR \leq SST. \quad (22)$$

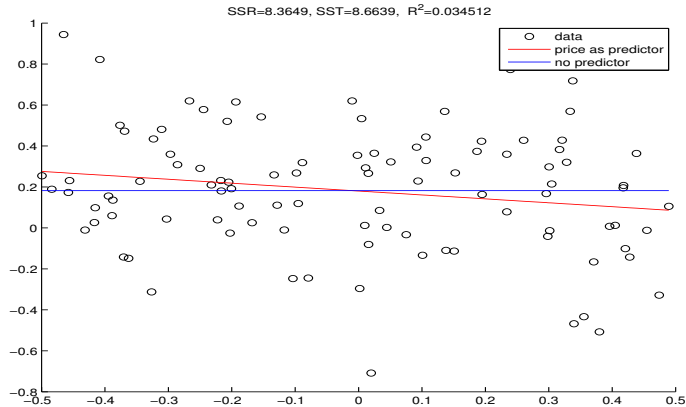
The coefficient of determination R^2 of the multiple regression model (18) is defined as:

$$R^2 = \frac{SST - SSR}{SST} = 1 - \frac{SSR}{SST}. \quad (23)$$

The coefficient of determination R^2 is a measure of goodness-of-fit:

- If $SSR \approx SST$, then there is little gained by including the predictors. R^2 is close to 0. The multiple regression model explains the variation in Y hardly better than the simple model (21).
- If $SSR \ll SST$, then much is gained by including all predictors. R^2 is close to 1. The multiple regression model explains the variation in Y much better than the simple model (21).





Proof of (22)

The following variance decomposition holds:

$$\text{SST} = \sum_{i=1}^N (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 = \sum_{i=1}^N \hat{\varepsilon}_i^2 + 2 \sum_{i=1}^N \hat{\varepsilon}_i (\hat{y}_i - \bar{y}) + \sum_{i=1}^N (\hat{y}_i - \bar{y})^2.$$

Using the algebraic properties (19) and (20) of the OLS residuals, we obtain:

$$\sum_{i=1}^N \hat{\varepsilon}_i (\hat{y}_i - \bar{y}) = \hat{\beta}_0 \sum_{i=1}^N \hat{\varepsilon}_i + \hat{\beta}_1 \sum_{i=1}^N \hat{\varepsilon}_i x_{1,i} + \dots + \hat{\beta}_K \sum_{i=1}^N \hat{\varepsilon}_i x_{K,i} - \bar{y} \sum_{i=1}^N \hat{\varepsilon}_i = 0.$$

Therefore:

$$\text{SST} = \text{SSR} + \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 \geq \text{SSR}.$$

1.1.3 Interpretation of the parameters

The parameter β_k is the expected absolute change of the response variable Y , if the predictor variable X_k is increased by 1, and all other predictor variables remain the same (ceteris paribus):

$$\begin{aligned} E(\Delta Y \mid \Delta X_k) &= E(Y \mid X_k = x + \Delta X_k) - E(Y \mid X_k = x) = \\ &\beta_0 + \beta_1 X_1 + \dots + \beta_k(x + \Delta X_k) + \dots + \beta_K X_K \\ &- (\beta_0 + \beta_1 X_1 + \dots + \beta_k x + \dots + \beta_K X_K) = \\ &\beta_k \Delta X_k. \end{aligned}$$

Interpretation of the parameters

The sign shows the direction of the expected change:

- If $\beta_k > 0$, then the change of X_k and Y goes into the same direction.
- If $\beta_k < 0$, then the change of X_k and Y goes into different directions.
- If $\beta_k = 0$, then a change in X_k has no influence on Y .

Question: Is the R^2 a good measure to use for model selection?

Question: Is the R^2 a good measure to use for model selection?

Question: What would we like to know about the OLS estimator next?

Question: Is the R^2 a good measure to use for model selection?

Question: What would we like to know about the OLS estimator next?

Question: What do we want to learn from the coefficient estimates?

Question: Is the R^2 a good measure to use for model selection?

Question: What would we like to know about the OLS estimator next?

Question: What do we want to learn from the coefficient estimates?

Revise: expectation of a random vector, variance covariance matrix, multivariate normal distribution, elementary properties on the: board

1. behavior under affine transformations,
2. behavior under transposition,
3. non-negative definiteness of a covariance matrix.
4. multivariate normality is preserved under affine transformations;

Statistical Properties of OLS Estimation

Econometric inference: learning from the data about the unknown parameter $\beta = (\beta_0, \dots, \beta_K)'$ in the regression model.

- Use the OLS estimator $\hat{\beta}$ to learn about the regression parameter.
- Is this estimator equal to the true value?
- How large is the difference between the OLS estimator and the true parameter?
- Is there a better estimator than the OLS estimator?

- Simulate data from a simple regression model with $\beta_0 = 0.2$ and $\beta_1 = -1.8$:

$$y_i = 0.2 - 1.8x_i + \varepsilon_i, \quad \varepsilon_i \sim \text{Normal}(0, \sigma^2) \quad (24)$$

- Run OLS estimation to obtain $(\hat{\beta}_0, \hat{\beta}_1)$ and compare the estimated values with the true values $\beta_0 = 0.2$ and $\beta_1 = -1.8$.
- Repeat this experiment several times

- Although we are estimating the *true model*, the OLS estimator differs from the true value.
- Many different data sets of size N may be generated by the same regression model due to the stochastic error term. The estimated parameters differ, as the sample mean, sample variance and correlation coefficient are different for each data set:

$$\hat{\beta}_1 = \frac{s_y}{s_x} r_{xy}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

What is the expected error of the OLS estimator? I

Obviously, the estimators are random variables. Hence it makes sense to study the statistical properties of OLS estimation:

- Are the OLS estimates unbiased, i.e., is the expected difference between the OLS estimator and the true parameter equal to 0?
- How precise are these parameter estimates, i.e., how large is the variance of the two estimators?
- Are the OLS coefficients correlated?
- How are the OLS coefficients distributed?

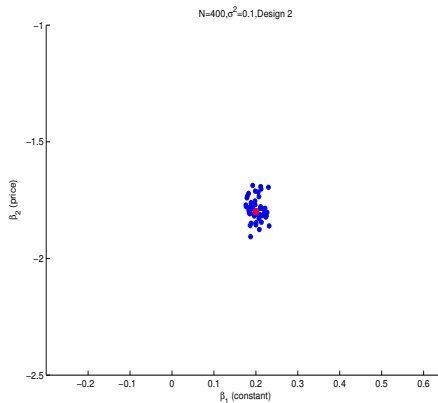
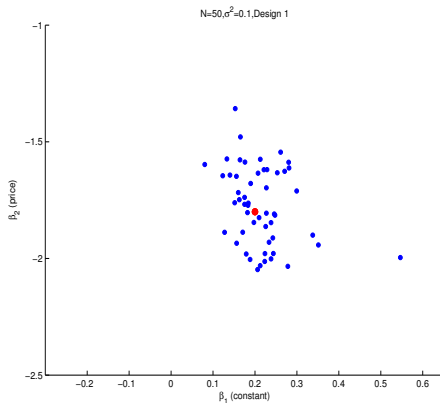
What is the expected error of the OLS estimator? II

For fixed values of $x_1, \dots, x_N$ the sampling properties of $\bar{y}$, s_y^2 and r_{xy} determine the estimation error. In general, the estimation error

- decreases within increasing number N of observations
- increases within increasing variance σ^2
- depends on the predictor variables through s_x^2

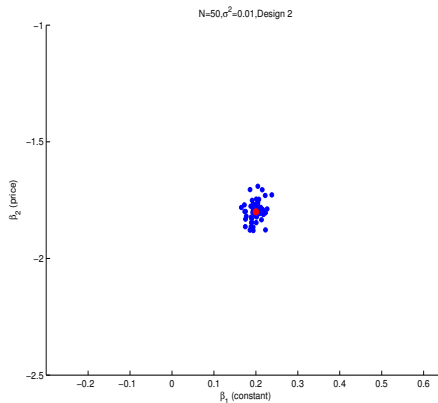
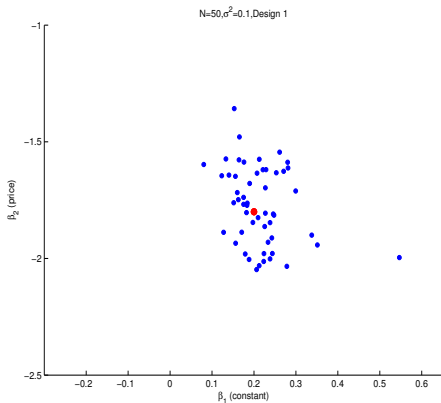
What is the expected error of the OLS estimator? III

$N = 50$ versus $N = 400$ ($\sigma^2 = 0.1$, Design 1)



What is the expected error of the OLS estimator? IV

$\sigma^2 = 0.1$ versus $\sigma^2 = 0.01$ ($N = 50$, Design 1)



Properties of estimator I

Assume that we have a set of N observations $y_i (i = 1, \dots, N)$ of a random variable Y , drawn independently from the same population with probability density $f(y_i, \theta)$:

- Unbiasedness: $\hat{\theta}$ is unbiased, if $E(\hat{\theta}) = \theta$. The bias is $E(\hat{\theta}) - \theta$.

All expectations are formed with respect to the sampling distribution of $\hat{\theta}$. The sampling distribution describes the probability distribution of $\hat{\theta}$ across possible samples.

- Mean squared error: the mean squared error (MSE) of $\hat{\theta}$ is the sum of the variance and the squared bias:

$$\text{MSE}(\hat{\theta}) = E((\hat{\theta} - \theta)^2) = V(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2.$$

- Efficiency: $\hat{\theta}$ is efficient if it is unbiased, and its sampling variance is lower than the variance of any other unbiased estimator $\hat{\theta}_A$: $V(\hat{\theta}) < V(\hat{\theta}_A)$.

Assumption AX: strict exogeneity:

the conditional expectation of ε conditional on all observations and variables in $\mathbf{X}$ is zero:

$$E(\varepsilon | \mathbf{X}) = \mathbf{0}, \quad E(\varepsilon_i | \mathbf{X}) = 0. \quad (25)$$

Knowing $\mathbf{X}$ does not give us any information about ε .

The term **endogeneity** (i.e., one or all explanatory variables are endogenous) is used, if **AX** does not hold (broadly speaking, if $\mathbf{X}$ and ε are correlated).

Strict exogeneity

Assumption **AX** implies:

- Conditional mean of Y , given $X_1, \dots, X_K$:

$$E(Y \mid X_1, \dots, X_K) = \beta_0 + \beta_1 X_1 + \dots + \beta_K X_K. \quad (26)$$

- Unconditional mean of the error in the population is 0:

$$E(E(\varepsilon \mid \mathbf{X})) = E(\varepsilon) = \mathbf{0}, \quad E(\varepsilon) = 0. \quad (27)$$

Regressors and disturbances are orthogonal, i.e., $\mathbf{X}$ and ε are uncorrelated:

$$\begin{aligned} E(x_{k,i}\varepsilon_j) &= E(E(x_{k,i}\varepsilon_j \mid \mathbf{X})) = E(x_{k,i}E(\varepsilon_j \mid x_{k,i})) = 0, \\ \forall i, j &= 1, \dots, N; \forall k = 1, \dots, K. \end{aligned} \tag{28}$$

These conditions imply that all $x_{k,i}$ and ε_j are uncorrelated, since

$$\text{Cov}(x_{k,i}, \varepsilon_j) = E(x'_{k,i}\varepsilon_j) - E(x'_{k,i})E(\varepsilon_j) = 0.$$

Assumption AH (homoskedasticity, assuming that AX holds):

$$V(\varepsilon_i | \mathbf{X}) = E(\varepsilon_i^2 | \mathbf{X}) - E(\varepsilon_i | \mathbf{X})^2 = E(\varepsilon_i^2 | \mathbf{X}) = \sigma^2. \quad (29)$$

If assumption (29) is violated, e.g., if $V(\varepsilon | \mathbf{X}) = \sigma^2 h(\mathbf{X})$, then we say that the error term is *heteroskedastic*.

It follows that

$$V(Y \mid X_1, \dots, X_K) = \sigma^2.$$

σ^2 is also the unconditional variance of ε , referred to as error variance:

$$V(\varepsilon) = E(\varepsilon^2) - (E(\varepsilon))^2 = \sigma^2.$$

Its square root σ is the standard deviation of the error.

Assumption AH :(Uncorrelatedness, assuming that AX holds):

$$\begin{aligned}\text{Cov}(\varepsilon_i, \varepsilon_j \mid \mathbf{X}) &= E(\varepsilon_i, \varepsilon_j \mid \mathbf{X}) - E(\varepsilon_i \mid \mathbf{X})E(\varepsilon_j \mid \mathbf{X}) \\ &= E(\varepsilon_i, \varepsilon_j \mid \mathbf{X}) = 0, \\ \text{Cov}(\boldsymbol{\varepsilon} \mid \mathbf{X}) &= E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \mathbf{X}) = \sigma^2 \mathbf{I},\end{aligned}\tag{30}$$

where $\mathbf{I}$ is the identity matrix.

This means that the errors are uncorrelated, regardless of the predictor variables $X_1, \dots, X_K$.